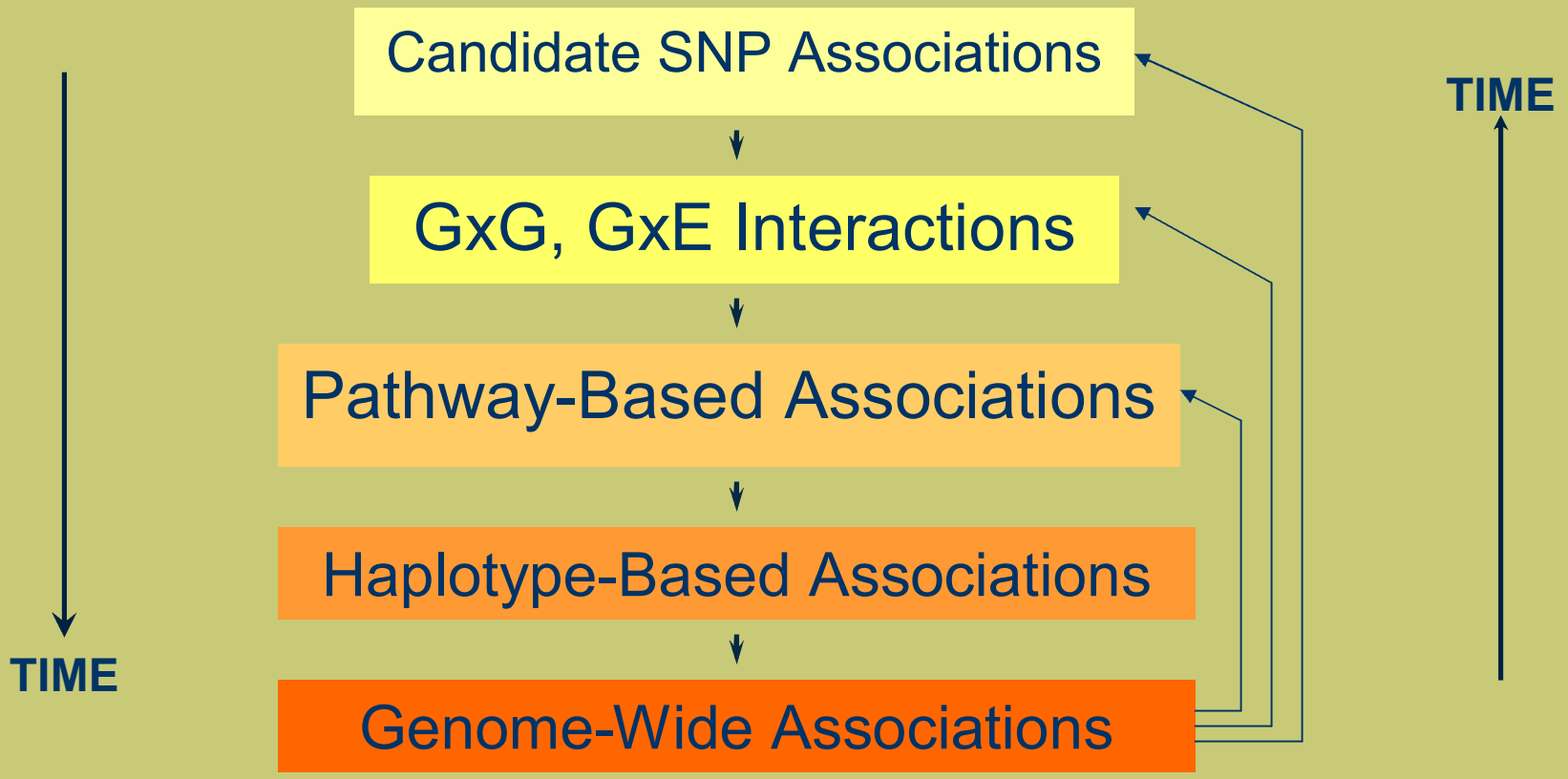
Associations, Biological Correlates, and Clinical Translation

Timothy R. Rebbeck, Ph.D.

University of Pennsylvania
School of Medicine

Evolution of Research



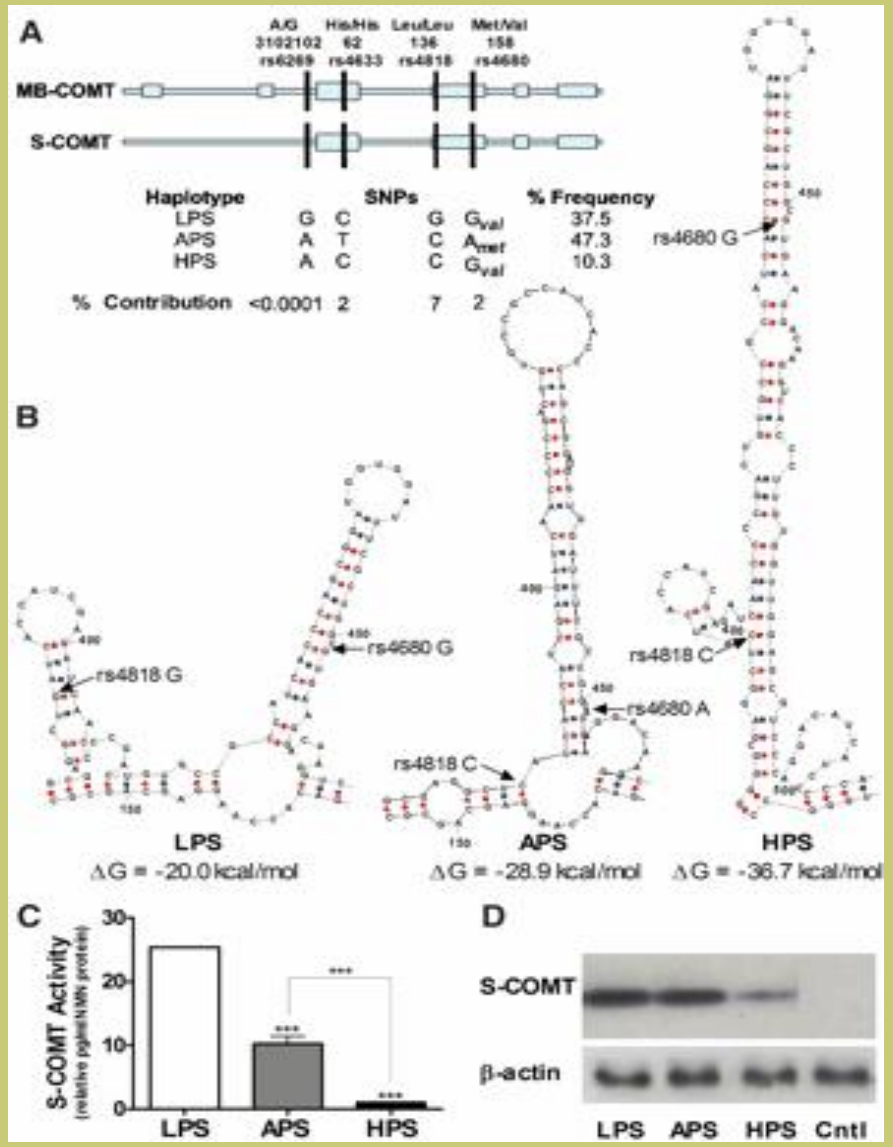
Goals

- *Predict Risk*:
  - Establish value-added information of genomic data (possibly in the absence of mechanism)
  - Example: *BRCA1* and breast cancer
- *Understand Context*:
  - Associations in relevant (biological) setting
  - Move from “Which model?” to “How many models?”
- *Define Biological Correlates*:
  - Provide biological insights
  - Explain/validate associations
  - Motivate risk prediction and prevention strategies

Using "Context" to Interpret Association Results

- SNP vs. Haplotype
- Population Structure
- Environmental Context

COMT Haplotypes, mRNA Secondary Structure and Enzymatic Activity



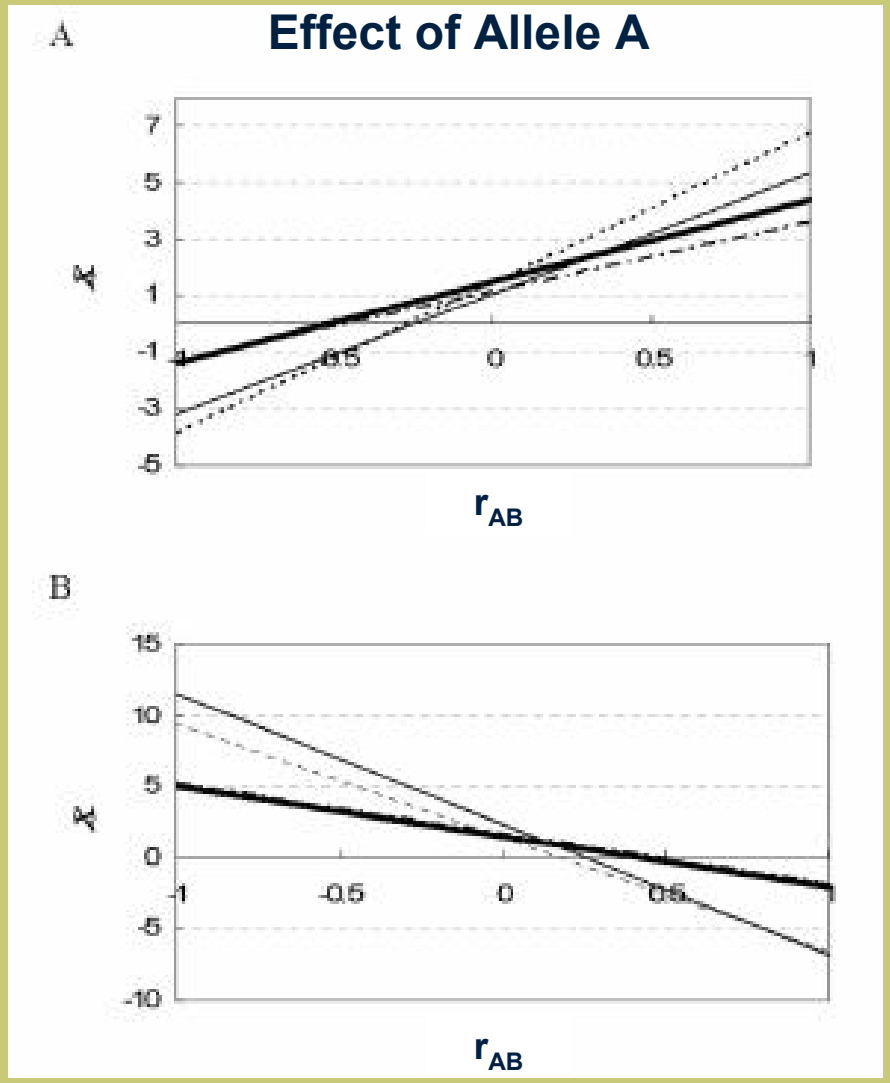
"Haplotypes divergent in synonymous changes exhibited the largest difference in COMT enzymatic activity"

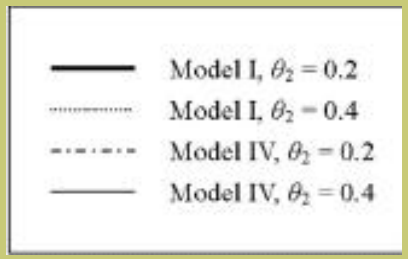
Nackley et al. *Science,* 2006

"Flip-Flop" Phenomenon

A+B are risk alleles

A is a risk allele, B is protective





Lin et al. *AJHG*, 2007

NAT2 and Bladder Cancer: Exposure Context

- NAT2 genotype associations have been inconsistent: opposite overall vs. in Benzidine exposed
- Exposure affects activation/metabolism by NAT2, and therefore risk:
  - Aryl-monoamines + NAT2 → Increased Risk
  - Aryl-diamines + NAT2 → Decreased Risk

Carreon et al. *Int J Cancer*, 2006

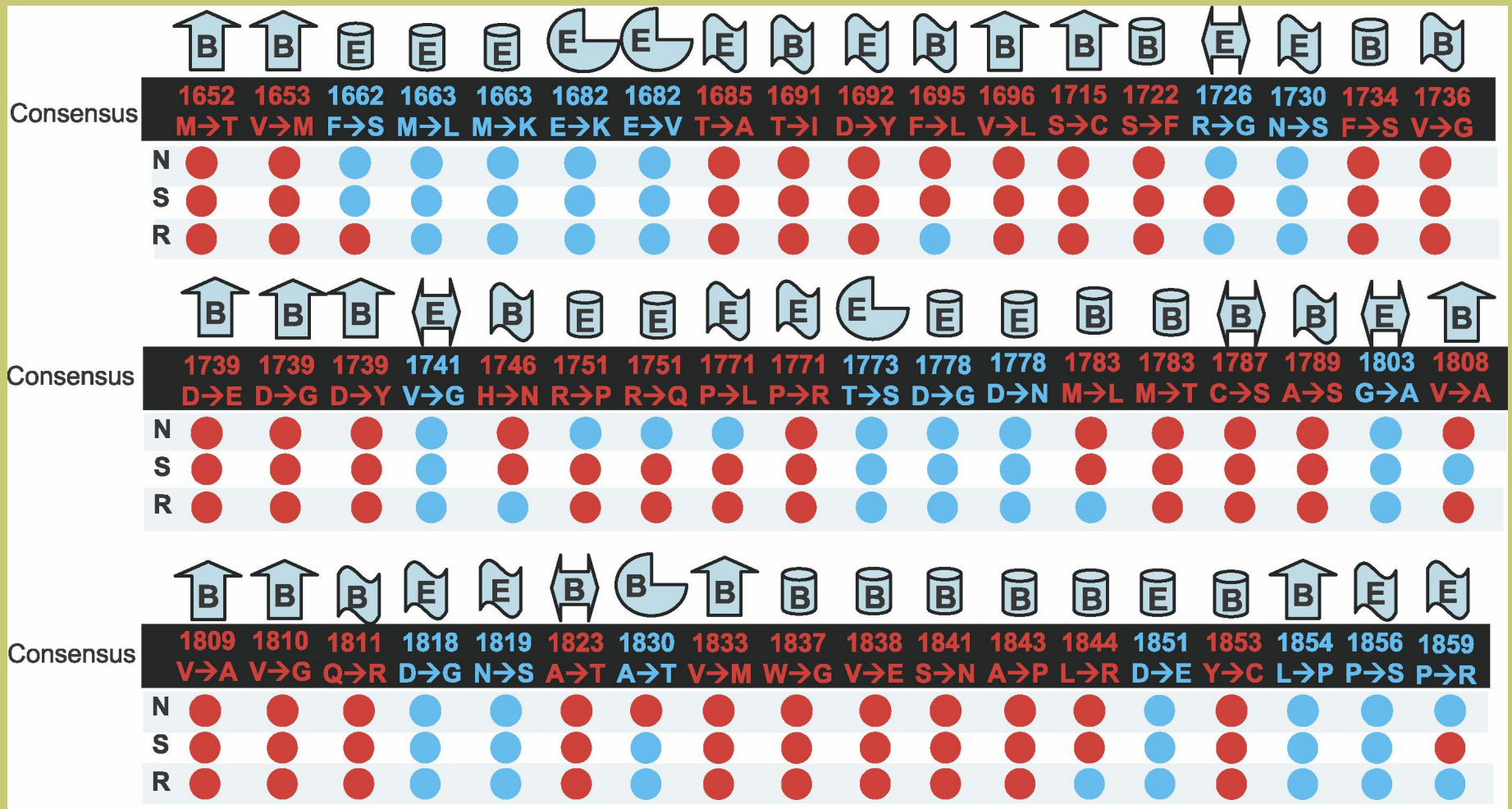
Goals

- *Predict Risk*:
  - Establish value-added information of genomic data (possibly in the absence of mechanism)
  - Example: *BRCA1* and breast cancer
- *Understand Context*:
  - Associations in relevant (biological) setting
  - Move from “Which model?” to “How many models?”
- ***Define Biological Correlates*:**
  - **Provide biological insights**
  - **Explain/validate associations**
  - **Motivate risk prediction and prevention strategies**

Sources of "Functional" Information

Evidence	Examples
Experimental	*In vivo, In vitro* assays
Nucleotide Sequence	Mutation Consequences
Evolutionary Conservation	Sequence Conservation (e.g., SIFT, CODDLE)
Population Genetics	Hardy-Weinberg, Linkage Disequilibrium
Exposures	Metabolism of relevant carcinogens
Epidemiology	Association consistency
Structural	Protein conformation (e.g., PolyPhen)

Computational Classifications of 54 Uncharacterized BRCA1 Variants

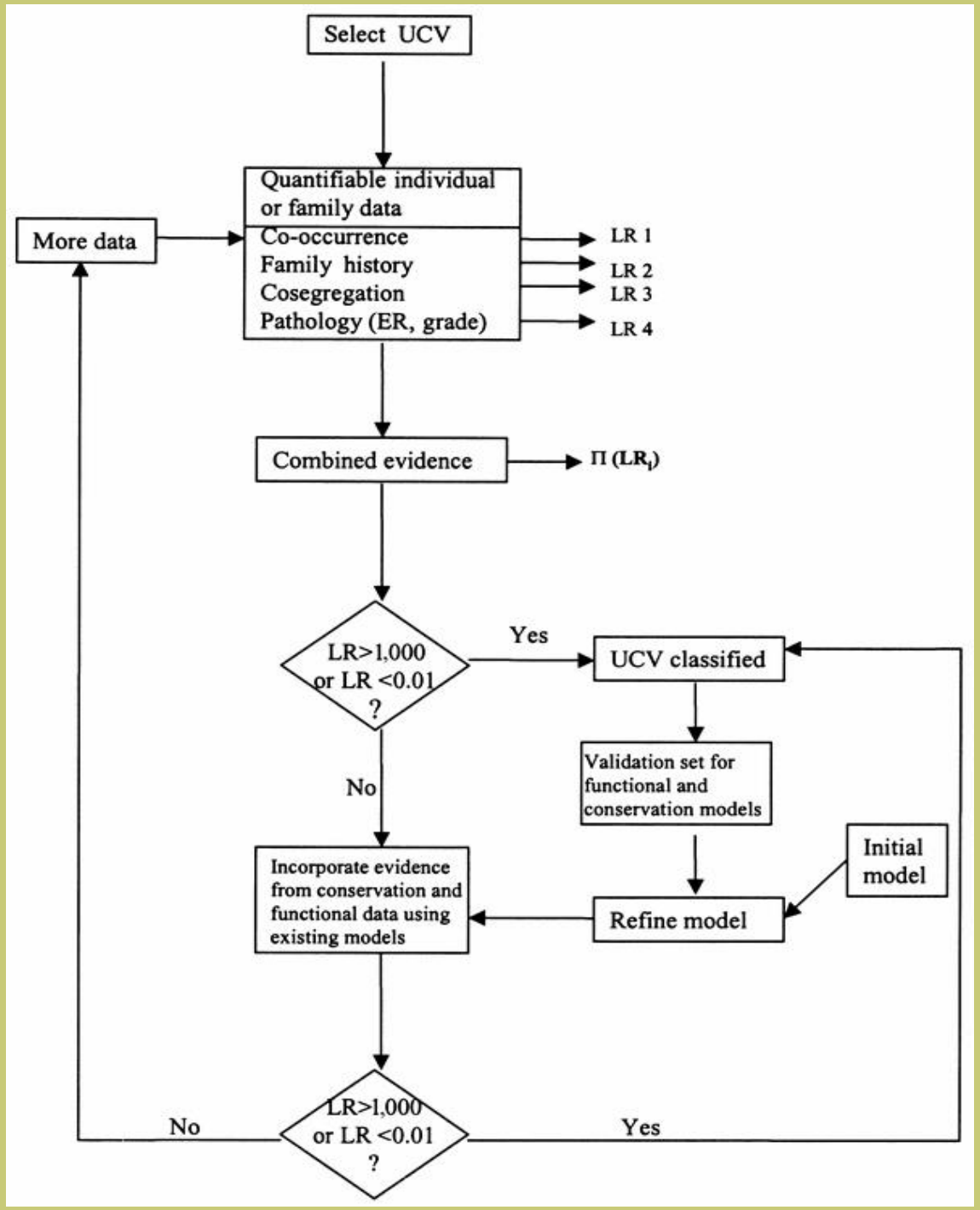


N=Naïve Bayes
R=Random Forest
S=Support Vector Machine

β-sheet
loop
bend
α-helix
H-bonded turn
B buried
E exposed
deleterious
neutral

Karchin et al.
***PLoS Comput Biol*, 2006**

Pathogenicity of BRCA1 Mutations Based on Pathology, Family History, Cosegregation, Co-occurrence with Disease



Goldgar et al.
***AJHG*, 2004**

Odds of Pathogenicity: Six SNPs in BRCA1 and BRCA2

Table 2

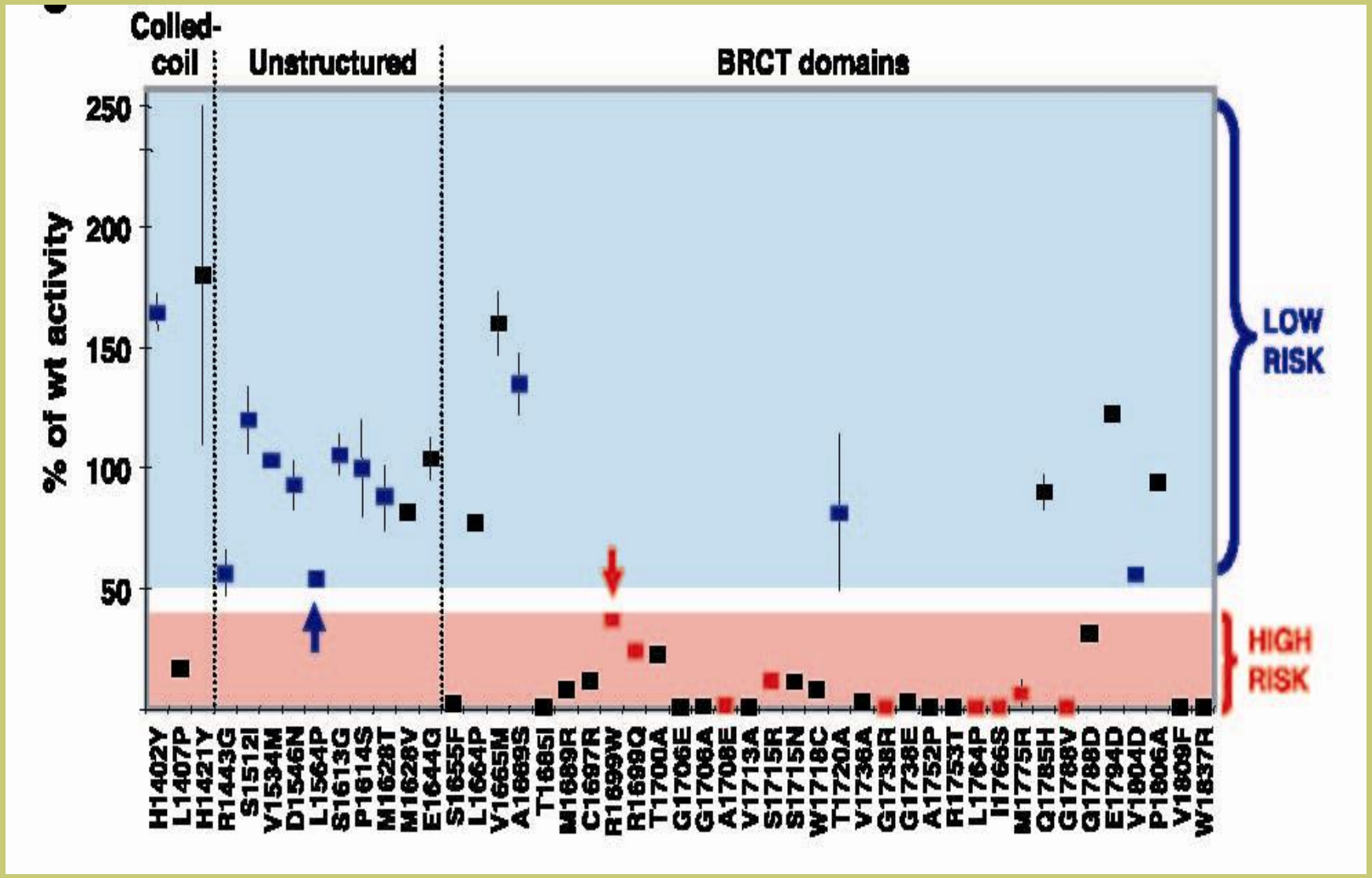
Odds in Favor of Each Variant Being Deleterious for the Six Variants Discussed in the Text, for Each Source of Information and Overall

	Odds in Favor of Causality for					
	BRCA1			*BRCA2*		
Data Source	C1787S	R1699Q	R841W	Y42C	P655R	D2723H
Co-occurrence	1.2	1.4	.028	8.9×10^{-11}	.007	2.0
Cosegregation	1,694	2.84	4×10^{-9}	6.7×10^{-7}	.48	13,731
GMS	1.5	.48	1.31	3.49	1.35	.98
Conservation	10.4	10.4	.006	.194[a]	.004[a]	5.0
Overall odds	31,692	20	8.7×10^{-13}	4×10^{-17}	.00002	134,563

[a] Deleted residue counted as a substitution.

Goldgar et al. *AJHG*, 2004

Transactivational Activity of BRCA1 Variants



Carvalho et al. *Cancer Research*, 2007

Problem: "Function" and Low Penetrance Genes

Rebbeck et al. *Nat Rev Genet* 2004

Effect of CYP3A4*1B Compared with CYP3A4*1A

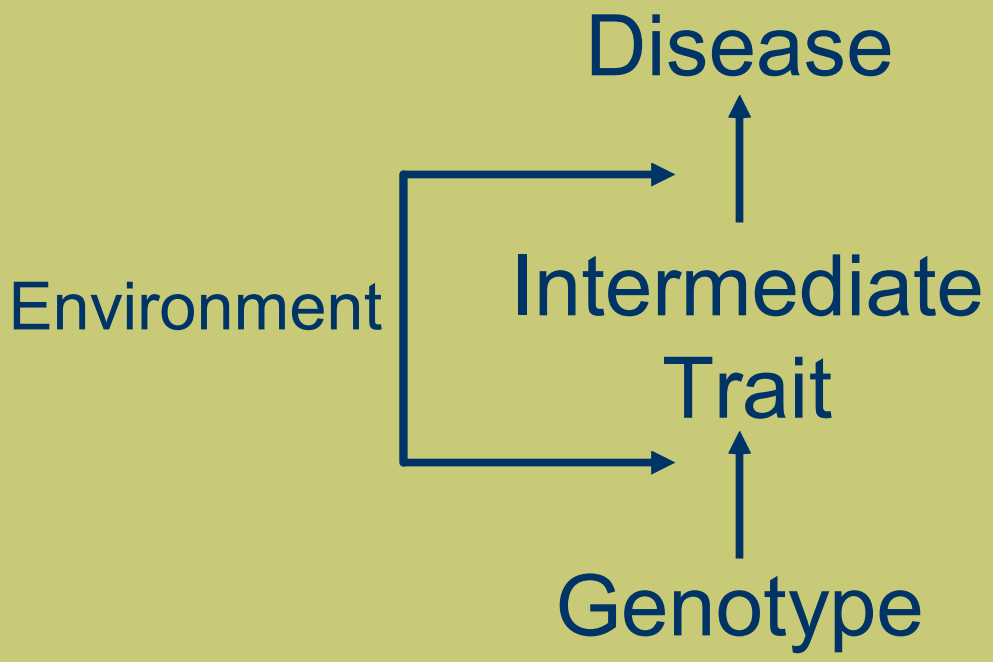
Westlind (Hepatocytes)	190% ↑ Testosterone oxidation
Amirimani (MCF7)	90% ↑ Luciferase Expression
Amirimani (HepG2)	40% ↑ Luciferase Expression
Ando (Hepatocytes)	40% ↑ Nifedipine Oxidase Activity
Ando (Hepatocytes)	110% ↑ CYP3A4 Protein Expression
Spurdle (HepG2)	No Change Luciferase Expression
Spurdle (HepG2+E)	20% ↑ Luciferase Expression
Amirimani (MCF7)	20% ↑ Luciferase Expression
Amirimani (HepG2)	40% ↑ Luciferase Expression
Amirimani (Hepatocytes)	20-90% ↑ Luciferase Expression
Hamzeiy (HepG2)	20-117% ↑ transcriptional activation*
Hamzeiy (HuH7)	75-147% ↑ transcriptional activation*

*Upon Xenobiotic Exposure

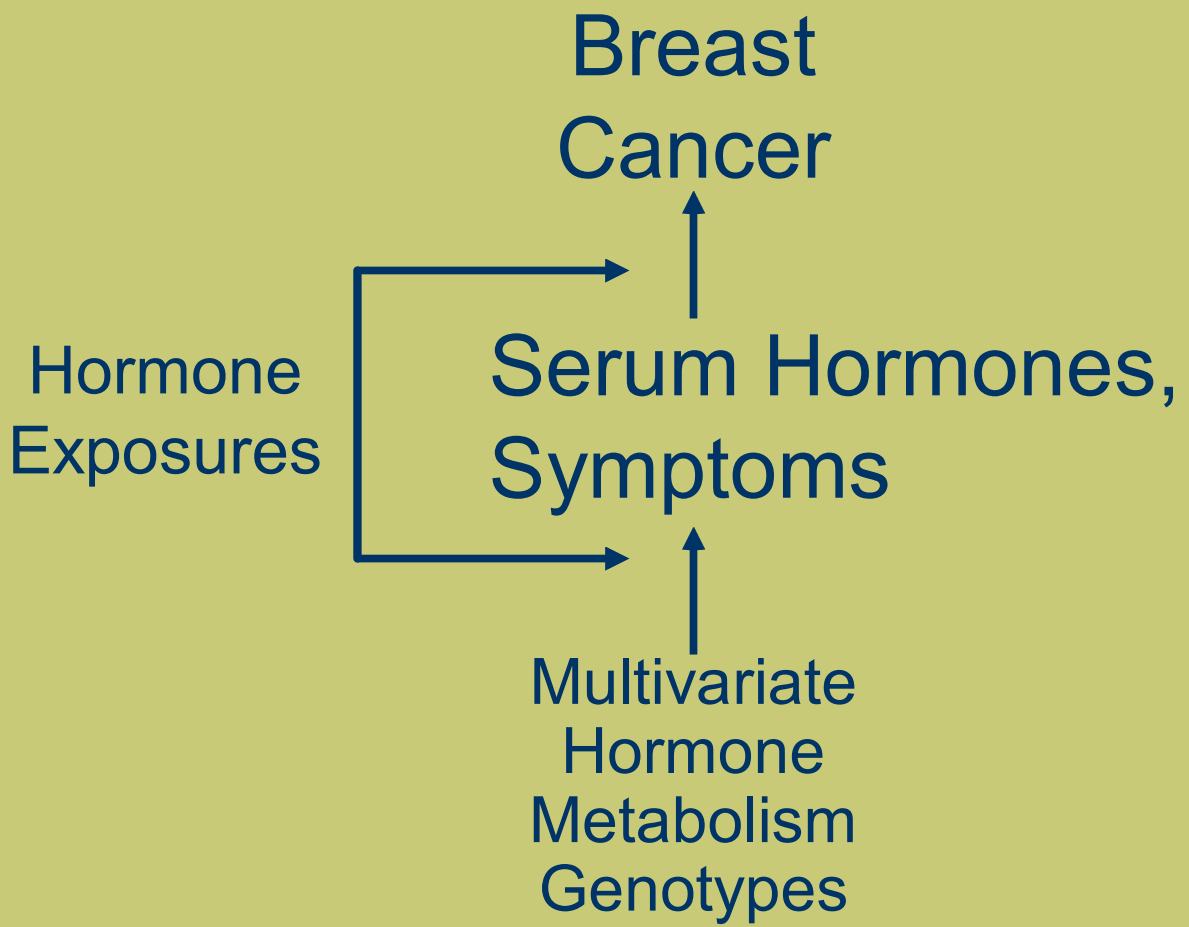
Goals

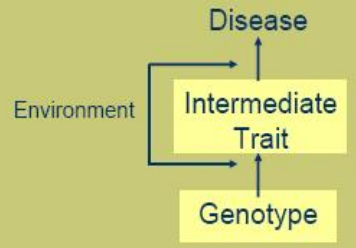
- *Predict Risk*:
  - Establish value-added information of genomic data (possibly in the absence of mechanism)
  - Example: *BRCA1* and breast cancer
- ***Understand Context*:**
  - **Associations in relevant (biological) setting**
  - **Move from "Which model?" to "How many models?"**
- ***Define Biological Correlates*:**
  - **Provide biological insights**
  - **Explain/validate associations**
  - **Motivate risk prediction and prevention strategies**

Defining Context & Biological Correlates for Low Penetrance Human Genes: Level Crossing Model



Breast Cancer Example: Level Crossing Model

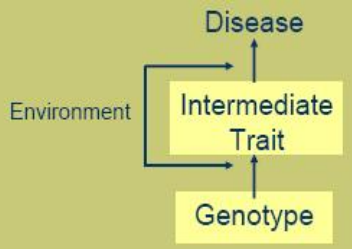




Estrogen Metabolism Genotypes and Postmenopausal Hormone Levels

	COMT	*CYP1A1*	*CYP1A2*	*CYP1B1*	*CYP17*	*CYP19*	*ESR1*	*HSD3B*
A-dione				O	OO	O		O
DHEA				O	O	OO		O
E1	O			O	OO	++		O
E2 Total	O+	O	O	OO	OO	++		O
E2 Follicular				O		O		
E2 Luteal				+		O		
E2 Free	O	+		O	O	++		
2-OH E1	O	O	O	O	O			
16α-OH E1	+	+	+	O	O			
FSH	O	O						
PG Total	OO	OO		OO	O	O		
PG Follicular				O		O		
PG Luteal				O		O		
SHBG	O	O	+	OO	O	O+		+
Testosterone	OO	OO		OO	O	OO		O

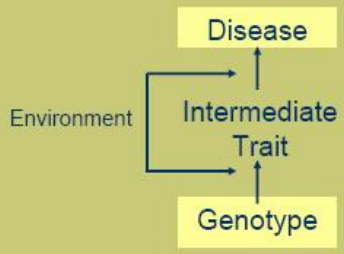
OO=VALIDATED NULL, O=NULL, 0+=MIXED, +=ASSOC., ++=VALIDATED ASSOC.



Estrogen Metabolism Genotypes and Menopausal Symptoms Longitudinal Cohort of 404 Women

Gene	Depression	Hot Flashes
COMT	1 18 (0.77-1.80)	0.95 (0.60-1.55)
CYP1A2	0.84 (0.52-1.34)	1.06 (0.63-1.79)
*CYP1B1*3*	0.64 (0.44-0.94)*	0.70 (0.48-1.04)
*CYP1B1*4*	0.80 (0.53-1.20)	1.00 (0.63-1.59)
CYP3A4	1.14 (0.72-1.81)	0.98 (0.60-1.63)
CYP19	1.10 (0.71-1.72)	1.41 (0.88-2.25)
*SULT1A1*2*	1.07 (0.76-1.50)	0.89 (0.60-1.31)
*SULT1A1*3*	1.15 (0.72-1.85)	0.91 (0.53-1.55)
SULT1E1 5'UTR	0.85 (0.55-1.32)	1.50 (0.96-2.34)
SULT1E1 3'UTR	0.90 (0.61-1.32)	0.99 (0.65-1.53)

*P<0.005 for interaction with menopausal status



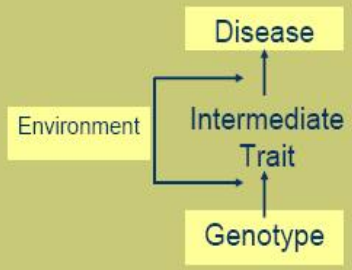
Hormone Metabolism Genotypes: Joint Effects

878 breast cancer cases and 1,409 age-matched controls

Table 3. Adjusted ORs, 95% CIs, and *P* values for pairwise genotype interactions and breast cancer case-control status

Gene	*COMT*	*CYP1A1*2C*	*CYP1A2*1F*	*CYP1B1*3*
COMT		14.27 (0.99-204.95) [0.051]	1.45 (0.27-7.78) [0.665]	2.05 (0.58-7.24) [0.263]
*CYP1A1*2C*	0.82 (0.28-2.36) [0.706]		1.65 (0.14-19.49) [0.692]	1.74 (0.39-4.83) [0.470]
*CYP1A2*1F*	1.37 (0.39-7.74) [0.628]	2.79 (0.45-17.5) [0.272]		1.49 (0.53-4.14) [0.449]
*CYP1B1*3*	0.95 (0.46-1.94) [0.880]	2.05 (0.63-6.66) [0.235]	0.68 (0.24-1.94) [0.477]	
*CYP1B1*4*	0.59 (0.31-1.12) [0.104]	0.96 (0.35-2.67) [0.940]	0.57 (0.21-1.54) [0.267]	*
CYP3A4	0.79 (0.25-2.45) [0.680]	3.73 (0.37-37.34) [0.262]	1.06 (0.24-4.77) [0.931]	0.78 (0.31-1.94) [0.590]
*SULT1A1*2*	1.91 (0.97-3.74) [0.061]	0.15 (0.05-0.42) [0.0004]†	0.50 (0.18-1.34) [0.167]†	0.87 (0.45-1.68) [0.673]
*SULT1A1*3*	0.35 (0.03-8.72) [0.387]			2.81 (0.29-27.49) [0.374]
SULT1E1	1.20 (0.48-2.96) [0.695]	1.81 (0.41-8.08) [0.440]	6.95 (1.72-28.00) [0.0064]†	0.87 (0.40-1.87) [0.715]

Rebbeck et al. *CEBP*, 2007



Progestin Metabolism Genotypes and Combined Estrogen/Progestin HRT

677 breast cancer cases and 905 age-matched controls

	CYP3A4			*PGR*		
Phenotype	**Never CHRT**	**<3 years CHRT**	**≥3 years CHRT**	**Never CHRT**	**<3 years CHRT**	**≥3 years CHRT**
Ductal	1.8 (0.8-3.8)	0.41 (0.1-1.6)	1.5 (0.6-3.6)	0.59 (0.3-1.2)	1.12 (0.3-5.0)	3.35 (1.1-10.0)
Lobular	2.5 (0.6-9.9)	NE	0.4 (0.1-4.1)	1.83 (0.7-5.0)	NE	1.08 (0.1-10.2)
PR +	2.4 (1.1-5.3)	0.65 (0.2-2.7)	0.7 (0.2-2.2)	0.59 (0.3-1.3)	1.52 (0.3-8.9)	3.82 (1.3- 11.6)
PR -	3.1 (1.1-8.7)	NE	0.7 (0.2-3.6)	1.58 (0.7-3.7)	NE	0.84 (0.1-7.7)
ER +	1.9 (0.9-4.3)	0.48 (0.1-1.9)	0.6 (0.2-1.8)	0.88 (0.4-1.7)	1.07 (0.2-6.1)	3.47 (1.2-10.2)
ER -	6.5 (2.0-20.7)	NE	1.2 (0.2-6.3)	1.09 (0.3-3.6)	NE	NE

NE: Not estimable

Rebbeck et al., *Am J Epi*, 2007

Questions

- How much and what kind of contextual and biological evidence do we need to make translational/clinical inferences about gene/variant pathogenicity?
- How to determine and disseminate the decision about functionality? Who will make the decisions?
- Will more complex models of functionality that include other data (e.g., disease characteristics, risk factors) improve understanding of pathogenicity?

Working Protocol

- Compute Sequence/Structure/Family analysis models
- Evaluate context of association:
  - Genome structure
  - Effect modification, joint effects
- Assess biological effects of genotype:
  - Phenotypic correlations
  - Expression and other genomic signatures
  - *In vivo* or *in vitro* function
  - DNA-protein and protein-protein interactions
- Reconcile computational, functional, and contextual data: “Venice Criteria” for functional data?
- Develop prediction models/tools
- Translate for risk assessment